

Máster en Programación avanzada en Python para Big Data, Hacking y Machine Learning

Programación Python para Machine Learning

LECCIÓN 12

Lección 12: Modelos no supervisados.

ÍNDICE

- ✓ Introducción
- ✓ Objetivos
- ✓ Principios teóricos de los métodos de clustering
- ✓ Implementación de un método de clustering particional: K-means.
- ✓ Implementación de un método de clustering jerárquico aglomerativo.
- ✓ Conclusiones

INTRODUCCIÓN



- ✓ Modelos no supervisionados.

✓ Datos no etiquetados.

- ✓ Cambio en la perspectiva del problema.

OBJETIVOS

Al finalizar esta lección serás capaz de:

1

Conocer los principios en los que se basan y la utilidad de los métodos no supervisados de Machine Learning

2

Analizar las distintas variantes de métodos no supervisados.

3

Dominar las técnicas de implementación de los métodos no supervisados en Python.

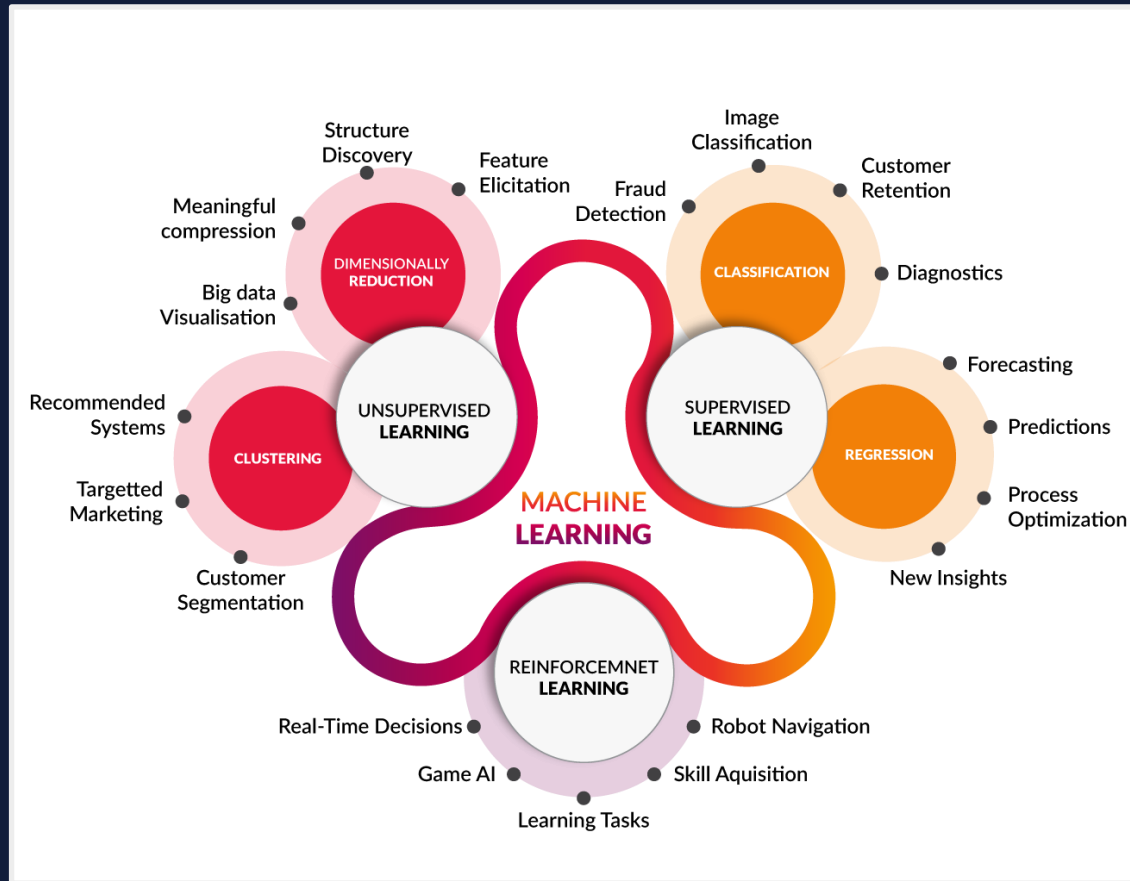
4

Interpretar los resultados de aplicar técnicas no supervisadas a conjuntos de datos.

MACHINE LEARNING NO SUPERVISADO

Modelos que se orientan a analizar la estructura intrínseca y común de los datos.

- ✓ Reducción de la dimensionalidad.
- ✓ Agrupación de patrones.



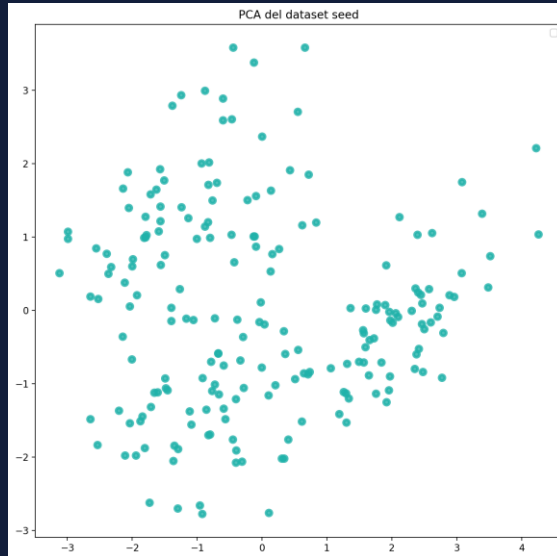
CLUSTERING

- ✓ Técnicas que buscan diferentes grupos dentro de las instancias de un conjunto de datos.
- ✓ Para ello, analizan las estructuras en los datos de manera que los elementos del mismo grupo sean lo más similares entre sí.
- ✓ Modelos más populares:
 - Métodos de clustering particional.
 - Métodos de clustering jerárquico.

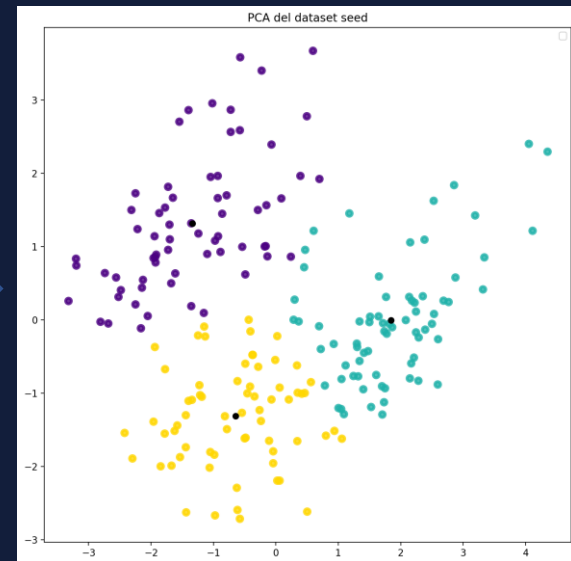
CLUSTERING PARTICIONAL

Tratan de obtener una partición de las instancias en grupos de tal forma que todas ellas acaben perteneciendo a alguno y que, a su vez, los clústeres sean lo más distantes posibles.

Los grupos se forman en base a la similitud entre instancias, cuantificando tal similitud en términos de distancia.



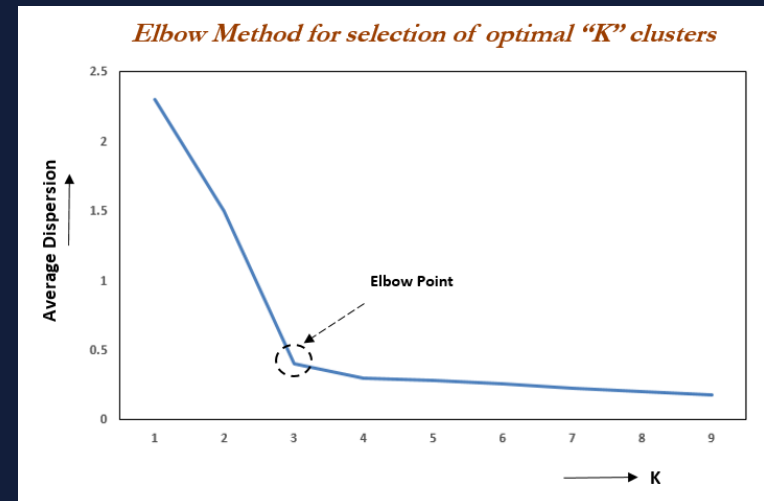
KMeans



CLUSTERING PARTICIONAL

1. Seleccionar aleatoriamente k instancias. Son considerados como centroides y se asigna el resto de los patrones al centroide que tenga más próximo.
2. Recalcular el nuevo centroide de todos los clústeres.
3. Reasignar cada instancia al clúster del centroide más cercano.
4. Repetir los pasos 2 y 3 hasta alcanzar un determinado criterio de parada.

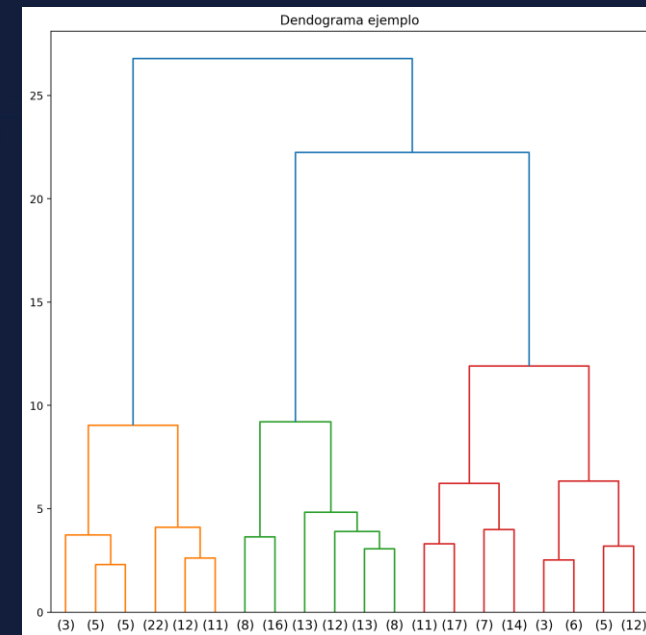
Distancia y k (número de clústeres)



CLUSTERIN JERÁRQUICO

El objetivo del clustering jerárquico, al igual que el clustering particional, es el de agrupar los patrones con características similares.

- ✓ Aglomerativo: enfoque ascendente
- ✓ Divisivo: enfoque descendente
- ✓ El algoritmo **aglomerativo**:
 1. Calcular la matriz de proximidad entre instancias.
 2. Considerar cada instancia como un clúster.
 3. Combinar los dos clústeres más cercanos.
 4. Actualizar la matriz de proximidad.
 5. Repetir los puntos 3 y 4 hasta que quede un solo clúster.



MUCHAS GRACIAS POR SU ATENCIÓN



jperez@grupomainjobs.com



Javier Pérez Rodríguez
www.linkedin.com/in/perezxavi



twitter.com/eiposgrados



facebook.com/eiposgrados



instagram.com/eiposgrados